Nonlinear Conjugate Gradient

usual CG

$$\phi(x) = \frac{1}{2} x^T A x - b^T x \quad \longleftrightarrow \quad Ax = b.$$

Can ϕ be a general convex fn.?

Can ϕ be a nonlinear fn ?

Yes!

Changes to be made

① Step length $\rightarrow$ Exact L.S. not possible

② Replace $r(x)$ by $\nabla f(x)$

Check new procedure $\rightarrow$ pick x_0

① Cacl $\nabla f(x_0) = r_0 = -p_0$

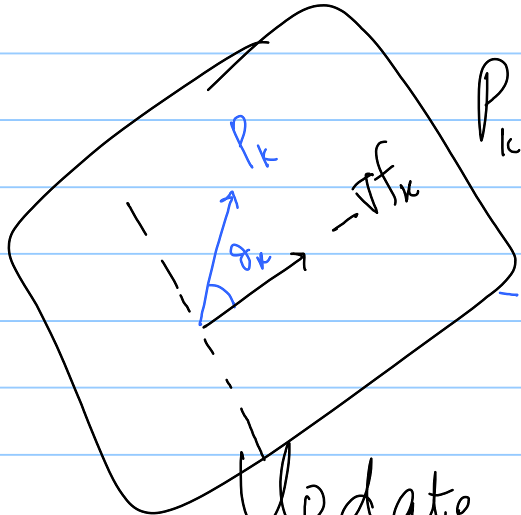
② Need $\alpha_0 \rightarrow$ inexact L.S. $\rightarrow x_1$

③ $$p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k$$

$$\beta_{k+1} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

Fletcher-Reeves NLCG.

Problem: A-conjugacy $\rightarrow$ GONE!



p_k needs to be a legitimate descent direction!

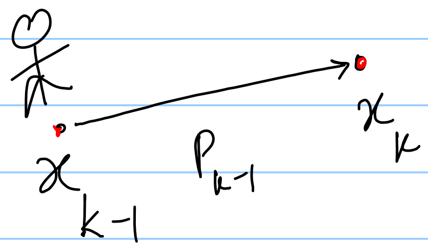
$$-\pi/2 < \theta_k < \pi/2$$

Update eqn: $p_k = -\nabla f_k + \beta_k p_{k-1}$

LM by ∇f_k^T

$$\nabla f_k^T p_k = \underbrace{-\|\nabla f_k\|^2}_{\text{great!}} + \beta_k \underbrace{\nabla f_k^T p_{k-1}}_{?}$$

Say that I did exact LS.



$$\phi(\alpha) = f(x_{k-1} + \alpha p_{k-1})$$

want best $\alpha \longrightarrow \alpha_{k-1}$

$$\Rightarrow \frac{d\phi(\alpha)}{d\alpha} = 0 = \underbrace{\nabla f(\overbrace{x_{k-1} + \alpha p_{k-1}}^{x_k})^T p_{k-1}}$$

$$0 = \nabla f_k^T p_{k-1}$$

Then $\nabla f_k^T p_k = -\|\nabla f_k\|^2$

$\Rightarrow$ If I do exact LS, & start with a DD

$\Rightarrow$ all p's are DD. ✓

In practical cases, exact LS ✗

The fix: Impose strong Wolf conditions

(w/o proof) with $0 < c_1 < c_2 < 1/2$.

then all dirs are descent dirs.

↳ Often $\theta_k \to \pi/2$. Polak-Ribiere

→ Set $\beta_k = 0$ (Reset) NLCG.

only one iter.

Newton & Newton-Like methods.

Ch 3, 6, 7 of NW

Motivation → Rate of conv quadratic

Price → Compute Hessian.

$$f(x,y) \to \quad \nabla f = \begin{pmatrix} \partial f/\partial x \\ \partial f/\partial y \end{pmatrix} \to \text{3 fn evals}$$

one sided

$$\left(\frac{f(x+h,y) - f(x,y)}{h} \right.$$

2 sided

$$\left. \frac{f(x+h,y) - f(x-h,y)}{2h} \right)$$

$\sim$ error: $f(x+h, y) \approx f(x,y) + h\frac{\partial f}{\partial x}(x,y) + O(h^2)$

error: $O(h^3)$

$$\nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x^2} \qquad \frac{\partial^2 f}{\partial x \partial y} \rightarrow \underbrace{\frac{\partial}{\partial x}\left(\underbrace{\frac{\partial f}{\partial y}}_{\text{FD in } y}\right)}_{\text{FD in } x}$$

It is a line search method:

$$x_{k+1} = x_k + \alpha_k p_k .$$

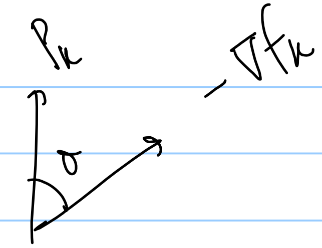
α_k ↓ inexact L.S.

$$p_k^N = -(\nabla^2 f_k)^{-1} \nabla f_k$$ ← Newton

$$p_k^{SD} = -(I) \nabla f_k$$

Hessian needed to be PD.

Show it leads to a legitimate D.D



$$0 < \cos\theta = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \, \|p_k\|} = \frac{\nabla f_k^T (\nabla^2 f_k)^{-1} \nabla f_k^T}{(\quad)}$$

$\forall x$.

A is P.D.

$z^T A z > 0$

$\Rightarrow$ if $(\nabla^2 f)^{-1}$ is P.D. the p_k^N is a legit D.D

$\nabla^2 f_k$ is P.D.

$$\nabla^2 f_k = Q \underbrace{\Lambda}_{>0} Q^T \qquad (\nabla^2 f_k)^{-1} = Q \underbrace{\Lambda^{-1}}_{>0} Q^T$$

$$p_k = -B_k^{-1} \nabla f_k$$

If $B_k = \nabla^2 f_k$ Newton

If B_k is an approx of $\nabla^2 f_k$,

quasi-Newton method.

— x —